Online Short-Term Forecast of System Heat Load in District Heating Networks

S. Grosswindhager[a], A. Voigt[b], M. Kozek[a]

[a]Institute of Mechanics and Mechatronics, Vienna University of Technology, Wiedner Hauptstr. 8, 1040 Vienna, Austria
[b]VOIGT+WIPP Engineers GmbH, Märzstr. 120, 1150 Vienna, Austria

Abstract

This paper presents an on-line short term forecasting approach for system heat load in district heating networks (DHN) using the popular Seasonal Autoregressive Integrated Moving Average (SARIMA) models in state space representation. The system heat load itself is a non-stationary random process composed of the individual consumer heat demands plus heat losses from pipes. Short term load forecasting is essential for effective operational production planning. It was found that the recurring pattern of the process based on half-hourly data are well described by a SARIMA(2,1,1)(0,1,1)48 model. The adequacy of the model was confirmed by standard regression diagnostics. Furthermore, the identified SARIMA model was incorporated into the state space framework where classical Kalman Recursion allows convenient calculation of on-line forecasting values. Moreover, exogenous effects such as weather effects are explicitly accounted for by decomposition of the original time series into an outdoor temperature dependent part and a social component part, where the latter was again modeled as SARIMA process. The relationship between system heat load and outdoor temperature may appropriately be expressed by a piece-wise linear function. Finally, the performance of the proposed model is validated on real data by calculating the mean absolute percentage error (MAPE) value for 48-steps-ahead (24h) estimates. The on-line performance for the basic and the temperature adapted model was assessed by computing rolling 24-steps ahead MAPE values for approx. 20 days of real data. In this work, the Kalman procedure is presented as an elegant approach for prediction of SARIMA processes in state space representation. Specifically, it is shown that the proposed methods are suitable for on-line short term forecasting of system heat load in district heating networks.

Keywords: Heat Load Forecast, SARIMA, State Space Models, Kalman Filter, District Heating Networks

1. Introduction

District heating systems are basically composed of heat production facilities, consumers stations and a distribution network. The former produces heat which is carried via a network of insulated pipes to each individual consumer where heat exchanger are used to transfer heat from the primary side of the distribution pipes to the secondary side of the building. It is well known that in such centralized heating production plants, reducing emissions and limiting pollution is more effectively accomplished than in local installations at the consumers. In Fig. 1 a simplified DHN is depicted with one heat source and two consumers. Typically, the plant operator only imposes some differential pressure Δp_p and supply temperature Ts_p on the DHN so that the supply temperatures and differential pressures at the consumer stations are maintained within the contractual limits [1]. Consumer installations then cause some return temperature and mass flow rates $\dot{m}$ in the network dependent on their space heating and tap-water heating demands. From an operational cost point of view, however, more efficient control strategies and optimal long and short term heat production planning are highly desirable.
A major research field and much attention is paid to forecasting the system heat load which can be formulated as

$$Q_{plant}(t) = Q_{loss} + \sum_i Q_{ci}(t + \tau_i(\dot{m})) \tag{1}$$

or directly the heat load of individual consumers [2], if sufficient consumer load time histories are available. Simple models are presented in [3, 4], where the heat load is split into a temperature dependent part and social behavior of the consumers. Sophisticated recurrent neural network model are proposed in [5]. Much work was also done on the

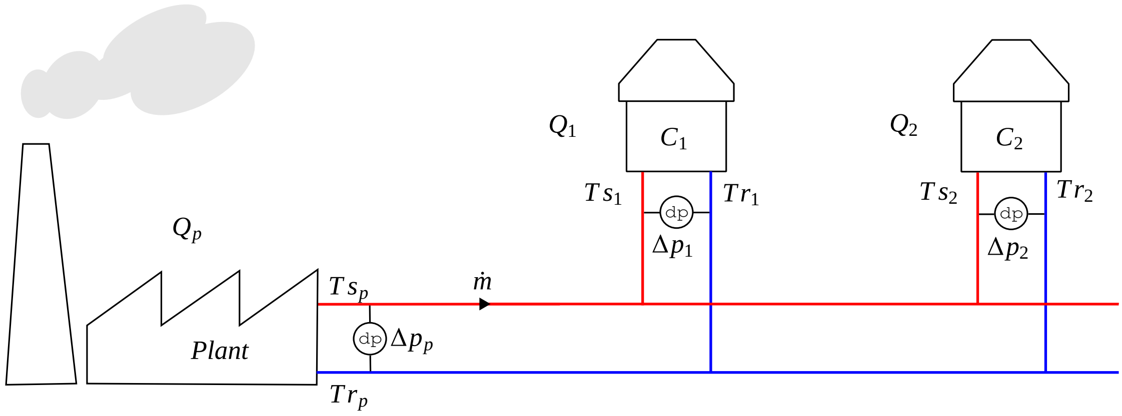


Figure 1: Simplified district heating network.

popular Box-Jenkins methodology [6], which is applied in [7] for heat load prediction or in [8] for electricity demand forecasting.
In this work a seasonal autoregressive integrated moving average (SARIMA) processes was proposed for modeling the system heat load. Consumer load forecasting were not treated, due to the highly stochastic nature of the consumer data, which would make it necessary to build several individual models. The system heat load is considered to be non-stationary random process mainly influenced by weather effects like outdoor temperature [4]. It can be argued, however, that for short term forecasting univariate SARIMA models can be sufficient because weather related variables tend to change in a smooth fashion and the essential dynamics is captured in the system load series itself [9]. Nevertheless, in addition to the univariate method the approach proposed [7] is followed, where exogenous influences such as temperature are accounted for by first filtering off these effects from the original load series and then identifying a SARIMA process for the the filtrated series. Furthermore, for the purpose of forecasting the estimated SARIMA model was incorporated into the versatile state space framework and classical Kalman recursion were applied. Such representations do not only have elegant forecasting abilities but also allows straightforward implementation together with other process models into operational planning tools.

2. Methodology

2.1. SARIMA Models

Seasonal autoregressive integrated moving average processes have been introduced in the literature to model time series with trends, seasonal pattern and short time correlations [10, 11, 6]. The SARIMA$(p, d, q) \times (P, D, Q)_s$ process representation for a series y_t can be stated in the following form

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \tag{2}$$

where s is the length of the periodicity (seasonality); ϵ_t is a white noise sequence;

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p, \qquad \Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} \cdots - \Phi_p B^{Ps} \tag{3}$$

denote the non-seasonal and seasonal autoregressive (AR) polynomial term of order p and P, respectively;

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 \cdots - \theta_q B^q, \qquad \Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} \cdots - \Theta_Q B^{Qs} \tag{4}$$

stand for the non-seasonal and seasonal moving average part (MA) a of order q and Q, respectively; finally $\nabla^d = (1-B)^d$ the non-seasonal differencing operator of order d is used to eliminate polynomial trends and the seasonal differencing operator $\nabla_s^D = (1-B^s)^D$ of order D is used to eliminate seasonal patterns. In the above equations B is the backshift operator, whose effect on a time series y_t can be summarized as $(B^d y)_t = y_{t-d}$. For the task of identifying appropriate order of the polynomials and differencing operators the sample autocorrelation (ACF) and partial autocorrelation functions (PACF) can be matched with the theoretical ones [6]. Competing SARIMA models can also be evaluated by considering information criteria such as AIC or BIC.

2.2. SARIMA in State Space Representation

There has been much interest in the embedding of (S)ARIMA processes into the framework of state-space models, for purposes of forecasting, as well as for specification and maximum likelihood estimation of parameters [6]. Following the notation of [12] a linear Gaussian state space model can be formulated as

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, H_t), & & \\ \alpha_{t+1} &= T_t\alpha_t + R_t\eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t), & t &= 1, \ldots, n, \\ & & \alpha_1 &\sim \mathcal{N}(a_1, P_1), & & \end{aligned} \tag{5}$$

where α_t is the unobserved state vector, y_t is the observation vector and the matrices Z_t, T_t, R_t, H_t and Q_t are referred to as the state space system matrices, which are assumed to be known. Furthermore, the error terms ϵ_t and η_t are assumed to be serially independent and independent of each other at all time points. The initial state vector is denoted as α_1 and has mean a_1 and variance matrix P_1 and is also assumed to be independent of the error terms at all time points. The former equation in (5) is referred to as the observation equation and has the structure of linear regression model. The latter equation is called the transition equation and has basically the form of a first order vector autoregressive model. State space representations of ARMA/ARIMA models are presented in [11, 6]. SARIMA models can be dealt with by constructing ARMA models for the stationary differenced series $y_t^* = (1 - B)^d(1 - B^s)^D y_t$ an placing the non-stationary variables such as y_{t-i} and $(1 - B)^d y_{t-i}$ in the state vector [13]. From (2) it is easy to see that y_t^* is a seasonal ARMA(p^*,q^*) model with order $p^* = p + SP$ and $q^* = q + SQ$. Let α_t^* denote the appropriately constructed state vector of this ARMA process. The complete state vector α_t has dimension $SD + d + m$ with $m = \max(p^*, q^* + 1)$ and, for instance, in the case of $d = 1$ and $D = 1$ can be written as

$$\alpha_t = (y_{t-1}, (1 - B)y_{t-1}, \ldots, (1 - L)y_{t-S}, \alpha_t^*)^T. \tag{6}$$

Due to the non-stationary terms in the state vector it is not clear how to populate the initial covariance P_1, because simply the observations $y_{1-i}, \quad i = 1 \ldots S$ are not available. To overcome this diffuse initialization of the state vector is applied, which is in detail discussed in [14, 15].

2.3. Forecasting State Space Models via Kalman Recursions

As already briefly outlined, the Kalman filtering allows an unified approach to prediction and estimation for all processes that can be given by state space representations. The classical Kalman recursions were introduced by Rudolph E. Kalman in 1960 [16]. The Kalman filter itself produces the minimum mean squared linear estimator for the state vector α_{t+1} based on the observations $Y_t = \{y_1, y_2, \ldots, y_t\}$, that is, $a_{t+1} = E(\alpha_{t+1}|Y_t)$ and the corresponding variance matrix of the estimator a_{t+1}, denoted as $P_{t+1} = Var(\alpha_{t+1}|Y_t)$ for $t = 1, \ldots, n$ [15]. Furthermore, given a_t and P_t the future value a_{t+1} can be obtained using the Kalman procedure as a linear function of the previous value and the forecast error of y_t, which means that the knowledge of the system can be updated each time a new observation comes in [12]. Due to the recursive structure it is also requisite to specify the initial state vector of the system as $\alpha_1 \sim \mathcal{N}(a_1, P_1)$. Note that due to the partially diffuse initial state vector of SARIMA models minor extensions to the classical Kalman filter equations are necessary [12, 17]. In general, l - step out of sample forecasts of future state values are recursively obtained as following $a_{t+l} = T_{t+l-1}a_{t+l-1}$ for $l = 1, 2, \ldots, L$ with covariance matrix generated from $P_{t+l} = T_{t+l-1}P_{t+l-1}T_{t+l-1}^T + R_{t+l-1}Q_{t+l-1}R_{t+l-1}^T$.

3. Empirical case-study Tannheim

3.1. SARIMA Modeling Process

Tannheim is located in Tyrol in Austria and is a typical tourist center with about 1100 inhabitants. Today, 84 building objects, mainly consisting of private houses, few hotels and some guest houses, are connected to this system.

The data for simulation consist of measurements of the system heat load between 18.05.2006 and 22.09.2010 with a resolution of 30 minutes. The entire dataset plus the snapshot of a specific week is depicted in Fig. 2. One can clearly observe the typical yearly cycles with high values occurring in winter months and low values in summer months and the daily heat load patterns with peaks in the morning hours and in the late afternoon. The inspection of ACF plot

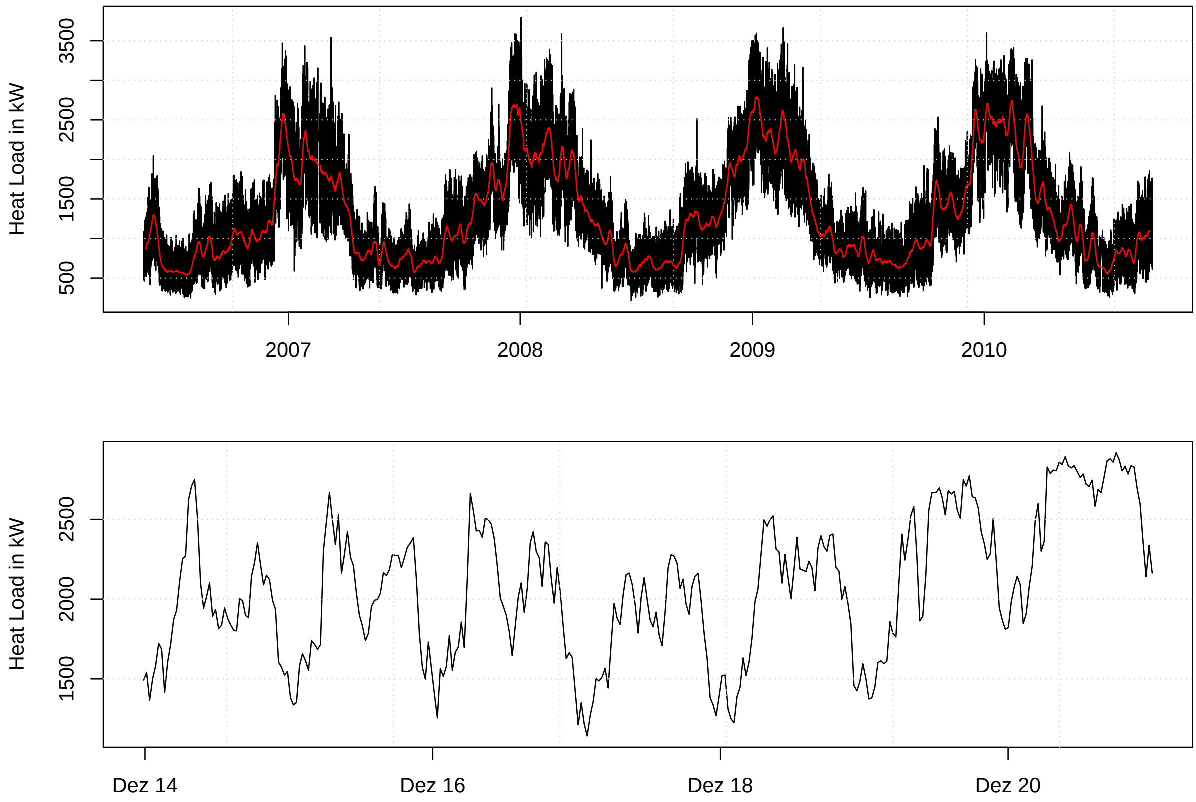
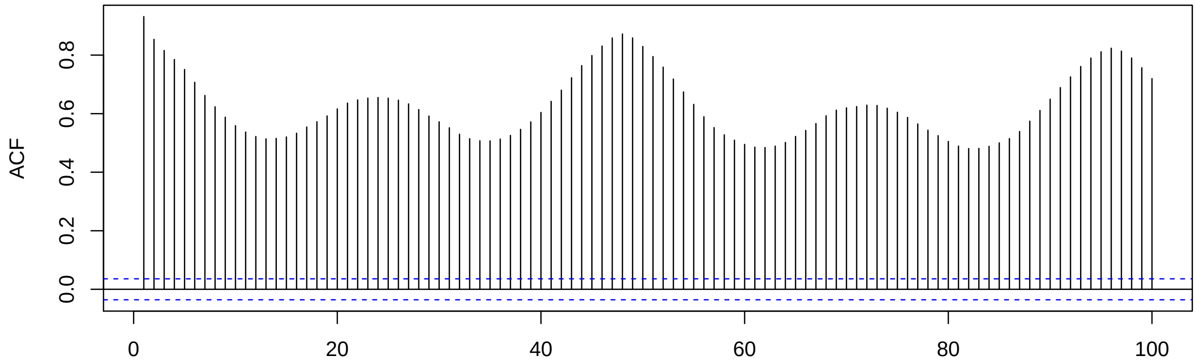


Figure 2: Half-hourly system heat load of DHN in Tannheim.

in Fig. 3 (top) provides vital structural information about the relationship between values of the series at different lags. Local maxima at lag multiples of 24 and 48 support the hypothesis of seasonal autocorrelations. The latter indicates high correlation between observations 24h apart, due to a sampling period of 30min. In order to make the series stationary seasonal differencing ($s = 48$) of degree one was applied. Moreover, to eliminate any upward and downward trends the data was double filtrated by also building first differences. Instead of additional detrending it is also possible to filter off the yearly temperature effects from the original time series, as proposed in [7].

The model selection procedure was performed in R 2.12.2 [18] based on the Bayesian information criterion. Due to computational issues 1000 samples (18.11. - 9.12.2007) were considered to be a sufficiently sized trainings dataset. Of course, the estimated parameter values depend on the chosen training data. Especially pronounced differences between summer heat load owing to tap-water heating and winter months where space-heating dominates can be expected. To overcome this one might use periodic models, as suggested in [13]. For this specific dataset the SARIMA$(2, 1, 1) \times (0, 1, 1)_{48}$ model achieved the lowest score. After having specified a tentative model the residuals should be checked for outliers, stationarity and normality. Figure 4 (top) reveals no systematic pattern of the residuals, but at least two suspiciously large std. residuals at observations 176 and 440 are observable. Such outliers can occur due to calendar effects (Christmas, Easter) and should be treated adequately with dummy variables [19]. Figure 4 (bottom) shows that at few lags the autocorrelations exceed two standard errors above zero ($\pm 2/\sqrt{n}$), with all values being very small. The Ljung-Box test with a p-value of 0.1939 (df=66) also does not reject null hypothesis, which states that the data is



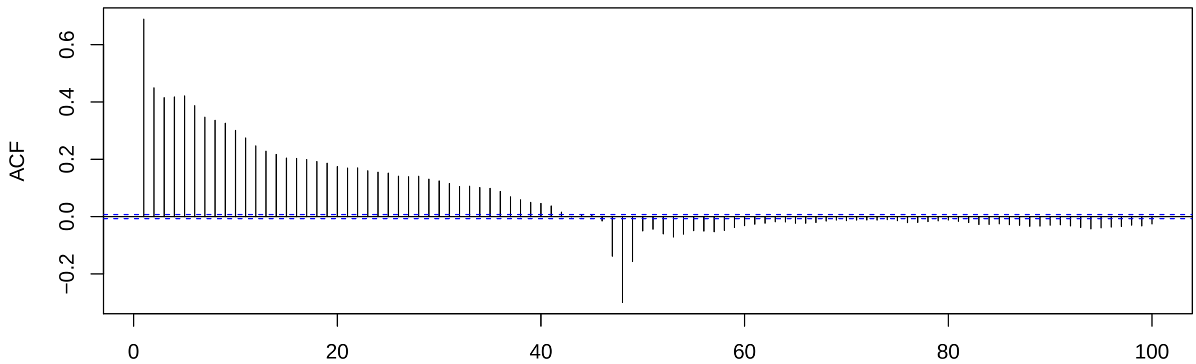


Figure 3: Autocorrelation function of system heat load before (top) and after seasonal differencing (bottom).

random. It seems the model have captured the essential dependencies in the series. Furthermore, the Shapiro-Wilk test produces a p-value of 0.08, hence the null hypothesis of normality of the residuals is not rejected at a 5% significance level. In the following the R- regression output is given with all coefficient estimates including the outlier effects being highly significant.

```
ARIMA(2,1,1)(0,1,1)[48]

Call: arima(x = data, order = c(2,1,1), seasonal=list(order = c(0,1,1), period = 48),io = c(176,440))

Coefficients:
         ar1      ar2      ma1     sma1     IO-176    IO-440
      0.2771  -0.3011  -0.4983  -0.8723  -222.8207  154.9367
s.e.  0.0861   0.0405   0.0926   0.0296    72.3473   71.7209

sigma^2 estimated as 9132:  log likelihood = -5726.21
AIC = 11464.42   AICc = 11464.54   BIC = 11498.43
```

The proposed model also satisfies the stationary and invertibility condition. (i.e. all roots of AR an MA characteristic polynomial lie outside the unit circle).

3.2. *Forecasting the Model using Kalman Recursions*

Next, the identified SARIMA model was incorporated into the state space framework for the purpose of forecasting. This work was performed in Matlab R2010a using the SSM Toolbox [17], where appropriate diffuse initialization methods for the state covariance matrix and Kalman filter recursions are already implemented. For the specific short term forecast horizons 12 to 24 hours were considered to be appropriate. Figure 5 shows the actual vs. the 24 hours

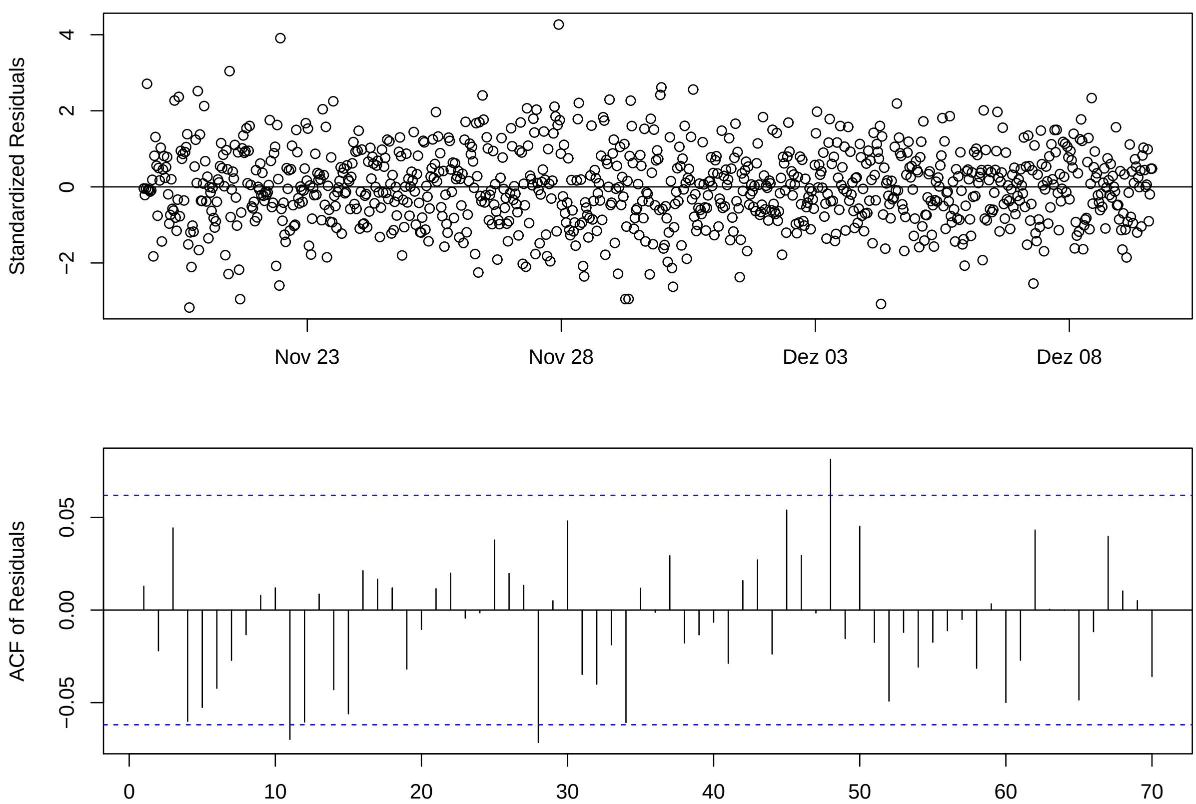


Figure 4: Residual diagnostic plots.

ahead prediction evaluated on observations between 16:47 12.02.2010 and 16:17 13.02.2010 for which the Mean Absolute Percentage Error (MAPE) was calculated to be 4.4%. Furthermore, exogenous effects are accounted for by constructing a temperature adjusted series through filtering off the temperature dependent part from the original time series (see [7]). The temperature dependency may be described by piecewise linear function $f_p(T)$ [4] robustly fitted to the data as depicted in Fig. 6. Here, a function with five segments is used, but the number can of course be chosen arbitrarily. It was found that the temperature adjusted series $y^+(t) = y(t) - f_p(T_t)$ can be adequately described by a SARIMA$(2,0,1) \times (0,1,1)_{48}$ process. Note that after forecasting of the filtered series, the outdoor temperature effects have to be included again, i.e. $\hat{y}_{t+k} = \hat{y}^+_{t+k} + f_p(T_{t+k})$. Hence, reliable and accurate temperature forecasts are necessary. To assess the on-line model performance rolling 12 hours-ahead prediction was performed and MAPE(24) were computed for the observations between 13.02. and 04.03.2011, as outlined in the following pseudo code:

Diffusive initialization of state vector with a_1 and P_1
for $t = 1$ **to** n **do**
    Given a_t and P_t
    Compute $(\hat{y}_{t+1}, \ldots \hat{y}_{t+k})$ and MAPE(k)
    Observe new observation y_{t+1}
    Update to a_{t+1} and P_{t+1} via Kalman filter
end for

The Boxplot plus mean with standard deviation of the MAPE(24) is shown in Fig. 7. As expected, the error is lower with inclusion of the outdoor temperature. However, it is necessary to mention that real temperature values were

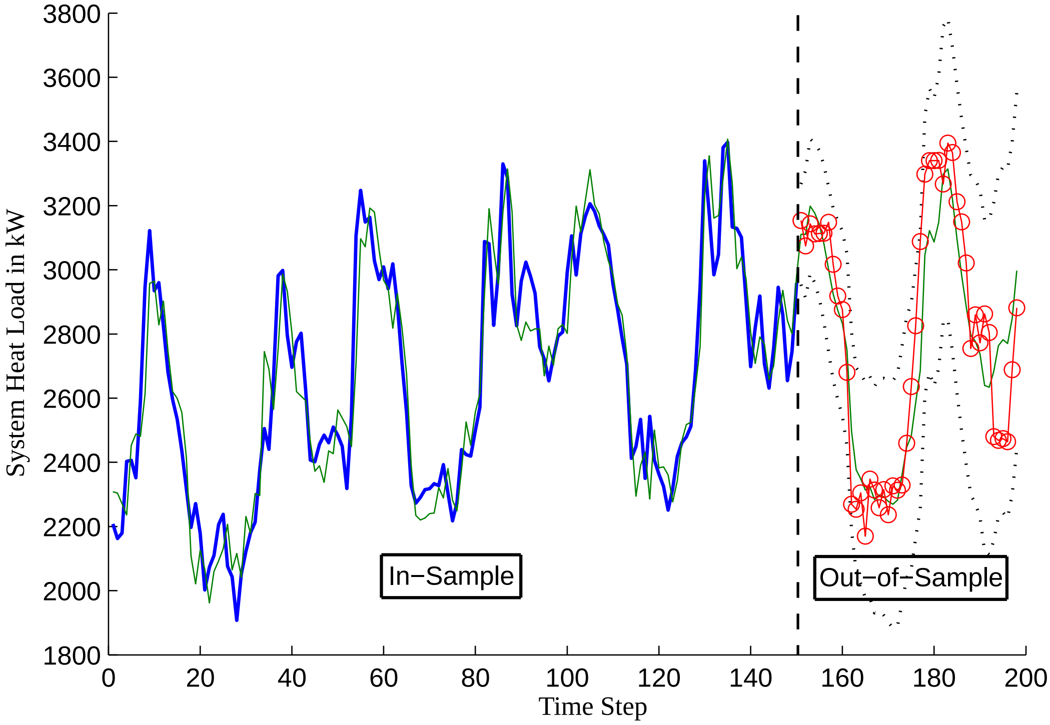


Figure 5: 48-step-ahead forecasts of system heat load from 16:47 12.02.2010 to 16:17 13.02.2010, including their 90% confidence interval.

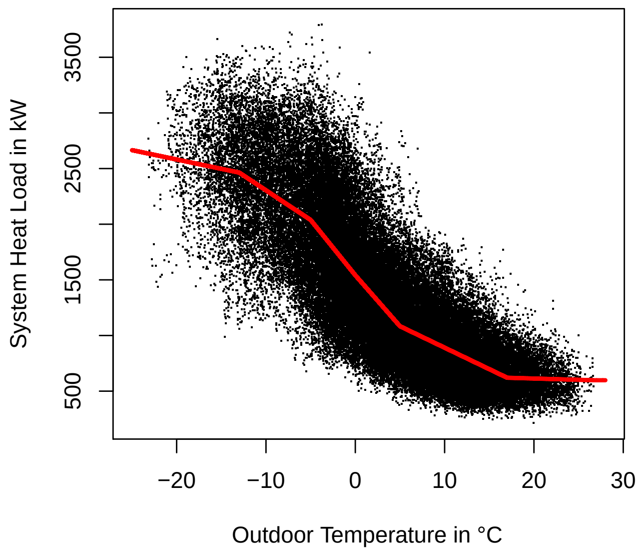
taken and not predicted ones. Some outliers at the basic forecast procedure without temperature correction could be explained, for instance, by calender effects or abrupt outdoor temperature changes. In practice, in such situations operator experience can be applied to supplement or override the proposed values from the standard forecasting system.

4. Conclusion

In this work it was shown that SARIMA models are appropriate for describing the system heat load of the Tannheim DNH. Furthermore, it was suggested to embed the identified model into the state space framework for forecasting purposes due to its ability of using Kalman recursions. The on-line performance was assessed by calculation of rolling MAPE(24) values. Outdoor temperature influences were taken into account by constructing a piecewise linear function, which lead to a significant improvement in terms of smaller prediction error. Alternative, more sophisticated forecasting methods such as time varying coefficient model approaches combined with structural models directly addressing trend, seasonal and cycle variations in the state space representation seem promising. Detailed analysis in this direction will be attempted in future publications.

5. Acknowledgment

This work was partly supported by Austrian Research Promotion Agency (FFG, No. 542534) and by VOIGT + WIPP Engineers.



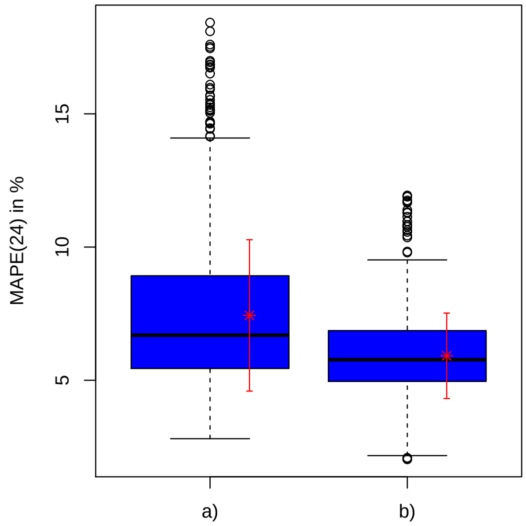


Figure 6: System heat load against outdoor temperature plus piecewise linear function describing the dependency.

Figure 7: Boxplot and $\overline{x} \pm \sigma$ bar of MAPE(24): a) Without inclusion of outdoor temperature; b) With inclusion of outdoor temperature.

References

[1] S. Meulen, Load management in district heating systems, Energy and Buildings 12 (1988) 179–189.

[2] M. Grzenda, B. Macukow, Heat consumption prediction with multiple hybrid models, Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living (2009) 1213–1221.

[3] G. Jónsson, A model for predicting the yearly load in district heating systems, Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy 216 (2002) 277–281.

[4] E. Dotzauer, Simple model for prediction of loads in district-heating systems, Applied Energy 73 (2002) 277–284.

[5] K. Kato, M. Sakawa, K. Ishimaru, S. Ushiro, T. Shibano, Heat load prediction through recurrent neural network in district heating and cooling systems, in: Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on, IEEE, pp. 1401–1406.

[6] G. Box, G. Jenkins, G. Reinsel, Time series analysis: forecasting and control, Wiley Series in Probability and Statistics, 4th edition, 2008.

[7] B. Chramcov, P. Dostál, J. Balátě, Forecast model of heat demand, in: The 29th Annual International Symposium on Forecasting, June 21-24, 2009, Hong Kong, pp. 1–9.

[8] Y. Kareem, A. Majeed, Monthly Peak-load Demand Forecasting for Sulaimany Governorate Using SARIMA., in: Transmission & Distribution Conference and Exposition: Latin America, 2006. TDC'06. IEEE/PES, IEEE, pp. 1–5.

[9] J. Taylor, P. McSharry, Short-term load forecasting methods: An evaluation based on european data, Power Systems, IEEE Transactions on 22 (2007) 2213–2219.

[10] P. Brockwell, R. Davis, Time series: theory and methods, Springer Verlag, 2009.

[11] P. Brockwell, R. Davis, Introduction to time series and forecasting, Springer Verlag, 2002.

[12] J. Durbin, S. Koopman, Time series analysis by state space methods, Oxford University Press, 2001.

[13] I. Hindrayanto, S. Koopman, M. Ooms, Exact maximum likelihood estimation for non-stationary periodic time series models, Computational Statistics & Data Analysis 54 (2010) 2641–2654.

[14] P. De Jong, The diffuse Kalman filter, The Annals of Statistics 19 (1991) 1073–1083.

[15] S. Koopman, Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models, Journal of the American Statistical Association 92 (1997).

[16] R. Kalman, A new approach to linear filtering and prediction problems, Journal of basic Engineering 82 (1960) 35–45.

[17] J. Peng, J. Aston, A MATLAB Software Implementation for Time Series Analysis by State Space Methods, in: Proceedings of the American Statistical Association Business and Economics Section [CD-ROM], American Statistical Association, 2006.

[18] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[19] J. Cryer, K. Chan, Time Series Analysis With Applications in R, Springer, New York, 2008.